# Evaluation of SMART-Seq Human BCR (with UMIs) for unbiased BCR repertoire profiling

**Shaveta Goyal*, Ritu Chandwani, Nick Tang, Jonathan Wang, Yue Yun, Robin Bombardi, Wenming Xiao & Andrew Farmer**

Takara Bio USA, Inc., 2560 Orchard Pkwy, San Jose, CA 95131 USA *Corresponding Author

## Abstract

Recent advancements in high-throughput next-generation sequencing (NGS) technologies have facilitated deep analysis of B-cell receptor (BCR) repertoires, providing profound insights into both healthy and diseased immune systems. However, accurate and sensitive NGS-based BCR profiling remains challenging because of the unparalleled diversity of BCR repertoires resulting from gene recombination, rearrangement, and somatic hypermutation events. In addition, NGS-based BCR profiling assays are prone to biases introduced through PCR amplification, use of multiplex primers, and sequencing platforms. In this study, we evaluated Takara Bio's BCR profiling solution, SMART-Seq® Human BCR (with UMIs), for such biases.

To assess the sensitivity, reproducibility, and specificity of SMART-Seq Human BCR (with UMIs), RNA from peripheral blood mononuclear cells (PBMC) was spiked with a mix of nine monoclonal B-cell lines, which have various known recombination events for heavy (IgG, IgA, IgM, IgD, IgE) and light chain (IgL, IgK) isotypes. The nine cell lines were mixed at different concentrations, ranging from 1% to 35% for a given individual cell line. Then 0.001%, 0.01%, 0.1%, 1%, or 10% of the cell line mixes were spiked into PBMC RNA. Using an RNA input of 25 ng, 100 ng, or 500 ng, libraries were prepared, sequenced on a NextSeq® 2000 platform, and analyzed using Cogent™ NGS Immune Profiler. To assess for bias in sequencing platform, a set of libraries were sequenced on three different Illumina® platforms: the MiSeq®, the NextSeq 2000, and the NextSeq 2000-XLEAP.

SMART-Seq Human BCR (with UMIs) effectively detected specific recombination events for heavy and light chain isotypes across the nine monoclonal B-cell lines. We demonstrated that the detection frequency for individual cell lines within the spike-in mixes correlates to the spike-in percentages or expression levels across all the technical replicates, suggesting that SMART-Seq Human BCR (with UMIs) is a specific and reproducible solution for BCR profiling. SMART-Seq Human BCR (with UMIs) also demonstrated strong linearity within the dilution range of 0.001% to 10%, with a limit of detection (LOD) of $\leq 10^{-6}$. Across sequencing platforms, we observed comparable LOD and reproducibility of sequenced BCR libraries. Our findings underscore the sensitivity and reproducibility of the SMART-Seq Human BCR (with UMIs) workflow and showcase its ability to detect low-abundance clonotypes without introducing bias.

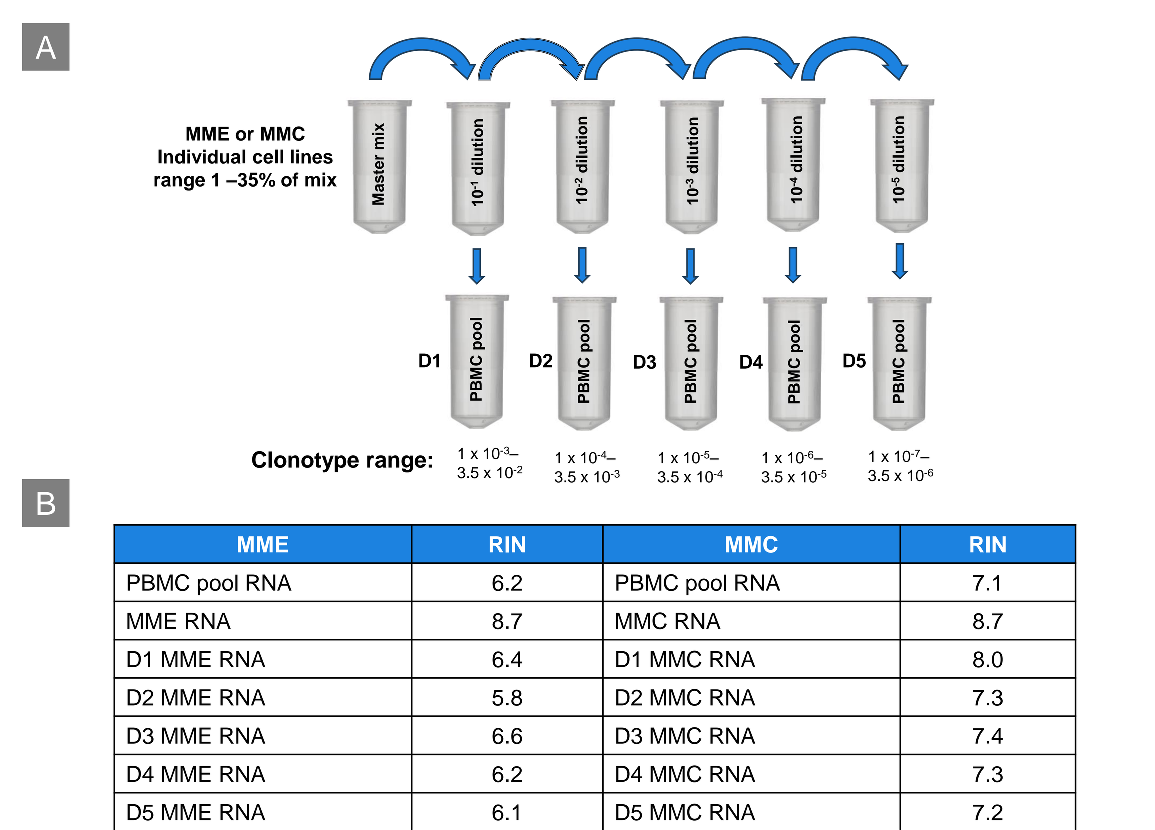
## Methods

BCR libraries in this study were generated from RNA samples shared by the US Food and Drug Administration (FDA). Briefly, spike-in master mixes were generated by pooling cells (Master Mix by Cells-MMC) from nine cell lines or extracting RNA from each cell line separately (Master Mix by Extraction-MME) and then pooling at a range of 1–35%. The cell line master mixes were serially diluted into a pool of PBMCs obtained from 13 healthy individuals to generate various spike-in concentrations with known clonotype fractions as low as 1 x 10-7 (Figure 1). Libraries were prepared with SMART-Seq Human BCR (with UMIs) (Takara Bio) using 25 ng, 100 ng, and 500 ng RNA inputs. The libraries were pooled for each concentration into a single pool and sequenced on NextSeq 2000. 25 ng libraries were also sequenced on Miseq and NextSeq 2000-XLEAP instruments.

Prior to producing the master mixes, cell line RNA for both MME and MMC batches was analyzed for expression levels using ddPCR assays specific for IgG and IgM. The assays were designed using the spike-in cell line sequence information (Figure 2). cDNA was generated from each of the cell lines with SMART-Seq Human BCR (with UMIs) using 10 ng of input RNA. 1 µl of resulting cDNA was diluted 100-fold, and the ddPCR assay was performed on a QX200 Droplet Digital PCR System (Bio-Rad).

Analysis of sequencing data was performed with Cogent Immune Profiler v1.7 (Takara Bio).

## 1 Generation of spike-in cell line mixes




| MME | RIN | MMC | RIN |
|---|---|---|---|
| PBMC pool RNA | 6.2 | PBMC pool RNA | 7.1 |
| MME RNA | 8.7 | MMC RNA | 8.7 |
| D1 MME RNA | 6.4 | D1 MMC RNA | 8.0 |
| D2 MME RNA | 5.8 | D2 MMC RNA | 7.3 |
| D3 MME RNA | 6.6 | D3 MMC RNA | 7.4 |
| D4 MME RNA | 6.2 | D4 MMC RNA | 7.3 |
| D5 MME RNA | 6.1 | D5 MMC RNA | 7.2 |

**Figure 1. Generation of spike-in cell line mixes. Panel A.** Cell line mix (MMC or MME) was serially diluted in the PBMC pool to generate samples with various clonotype fractions, with the lowest being $10^{-7}$ **Panel B.** RNA quality of cell line mix (MME or MMC), the PBMC pool, and the PBMC pools with spike-ins of serial dilutions of cell line mix.

## 2 ddPCR assay design

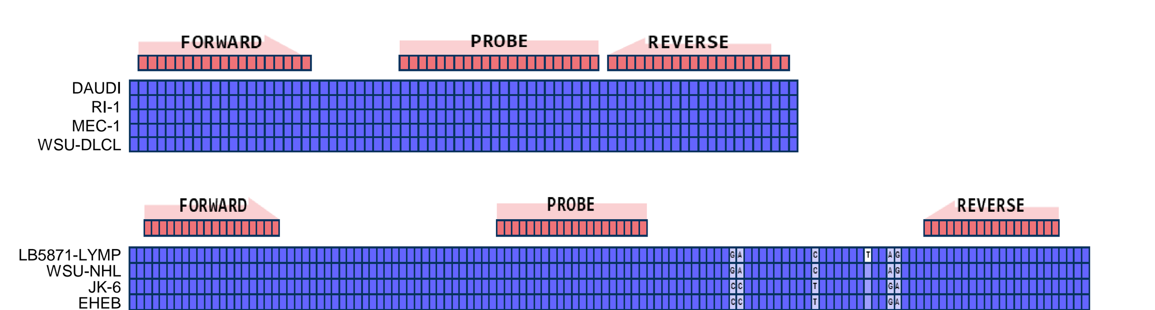



**Figure 2. ddPCR assays were designed based on the IgM or IgG sequences specific to the cell lines used in the study.** Names of individual cell lines are shown on the left. ddPCR assays for DAUDI, RI-1, MEC-1, WHU-DLCL, and OCI-LY1 cell lines target IgM. ddPCR assays for LB5871-LYMP, WHU-NHL, JK-6, and EHEB cell lines target IgG.

## 3 Clonotype identification across RNA input amounts

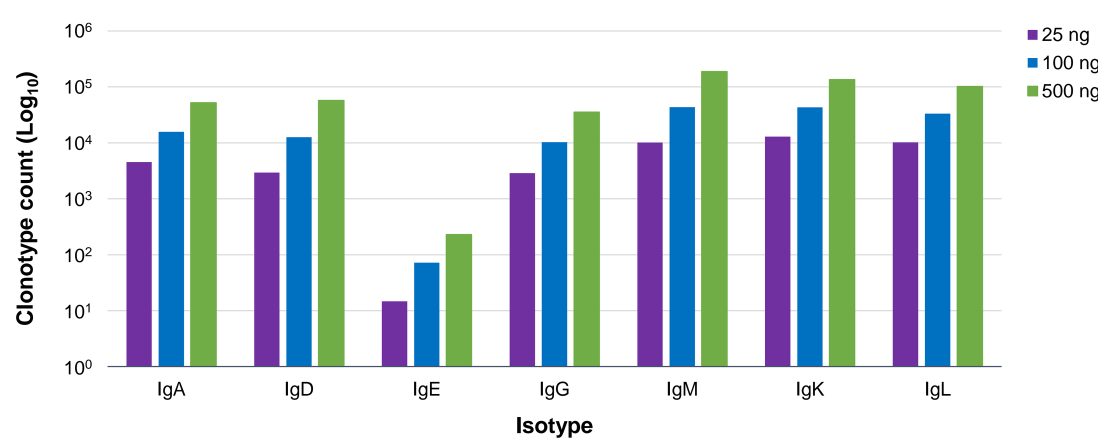



**Figure 3. Clonotype counts across Isotypes from libraries prepared with SMART-Seq Human BCR (with UMIs) using PBMC RNA at various inputs.** An exponential increase in clonotype count across isotypes is observed when increasing the amount of PMBC RNA input into the SMART-Seq Human BCR (with UMIs) workflow from 25 ng to 500 ng. The read depths for the 25 ng, 100 ng, and 500 ng RNA input libraries were 2.5 x $10^6$, 1 x $10^7$, and 1.5 x $10^7$ reads, respectively.

## 4 Reproducibility between technical replicates

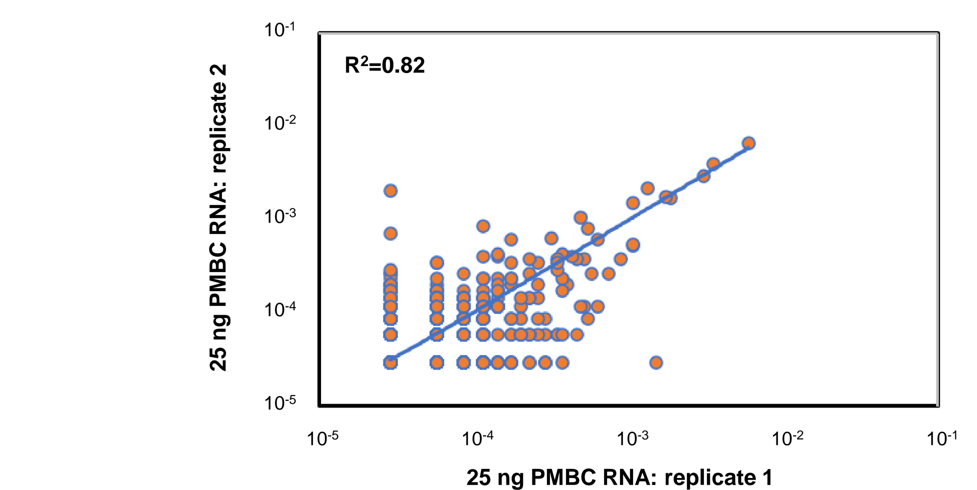



**Figure 4. Reproducibility of clonotypes identified from libraries prepared with SMART-Seq Human BCR (with UMIs) using 25 ng of PBMC RNA.** Excellent reproducibility between technical replicates is observed, with an $R^2$ value of 0.82

## 5 ddPCR assay: expression abundance

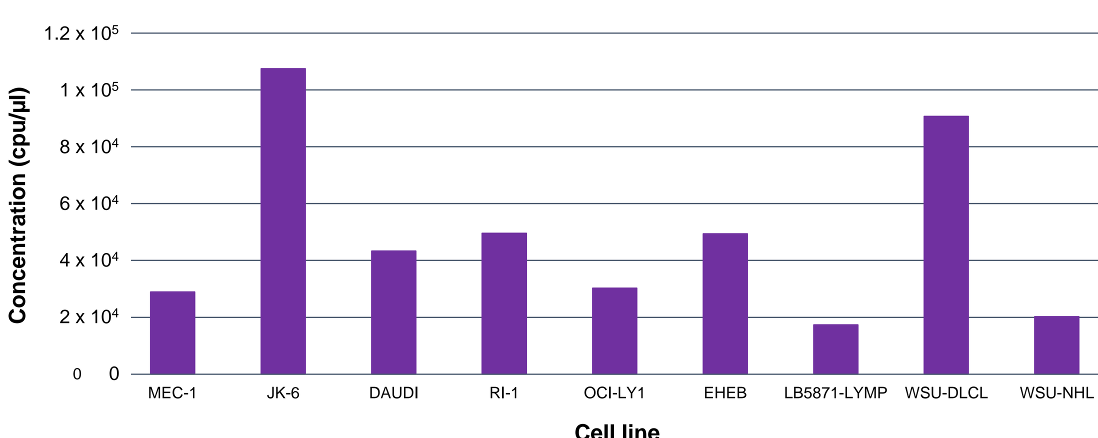



**Figure 5. Abundance of individual spike-in cell lines in the MME cell mix as determined by the ddPCR assay.** ddPCR results show that JK-6 is the most abundant, followed by WSU-DLCL

## 6 Expected vs. observed cell line proportions

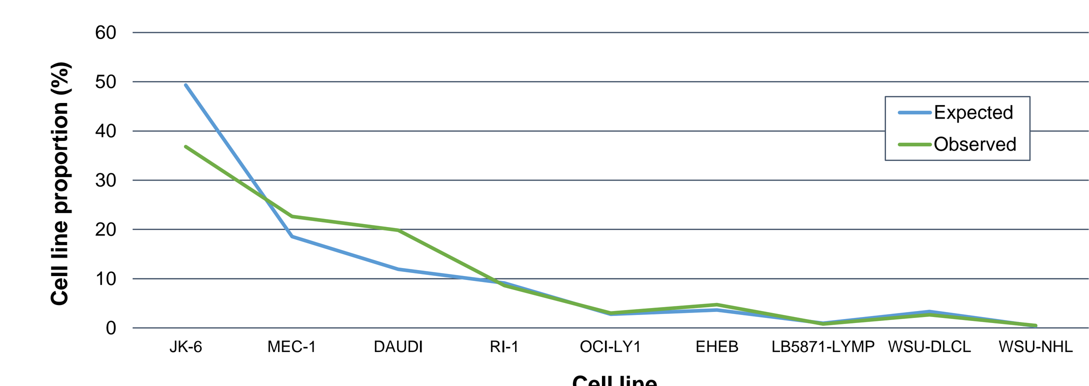



**Figure 6. Proportions of individual spike-in cell lines in the MME cell mix as determined by ddPCR.** Expected proportions were normalized by factoring in the cell line spike-in percentage for each cell line. After normalization, the abundance determined by ddPCR correlates to the NGS data. JK-6 is the most abundant cell line followed by MEC-1. WSU-NHL is the least abundant cell line, which correlates to the NGS data.

## 7 Linearity of detection for individual cell lines in spike-in cell line mixes

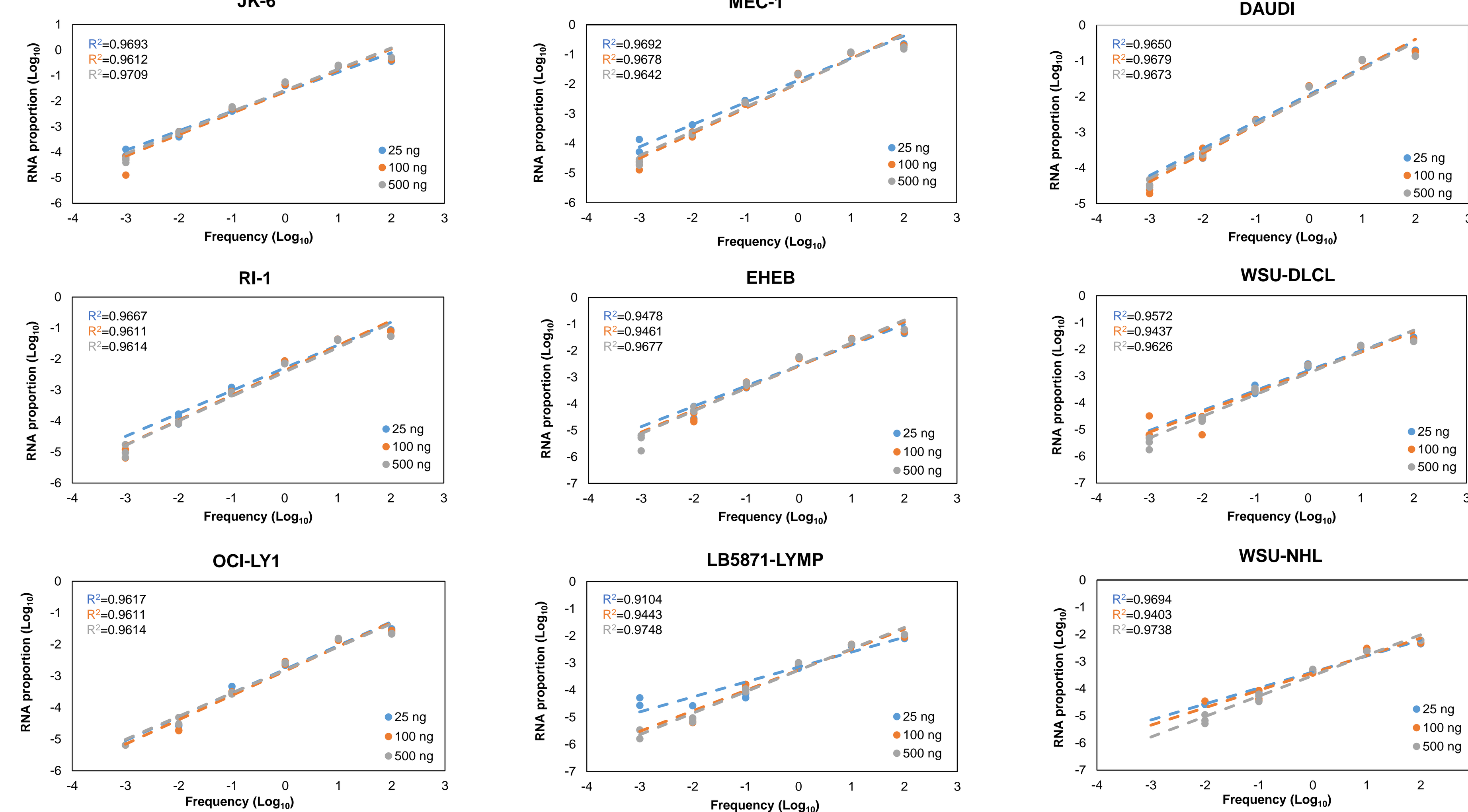



**Figure 7. Linearity of detection for individual spike-in cell lines in the MME cell mix.** Low abundant clones are successfully detected in libraries produced using SMART-Seq Human BCR (with UMIs) and there is excellent correlation ($R^2$>0.9) between spike-in RNA proportions and detected frequencies for each cell line.

## 8 Sensitivity and accuracy

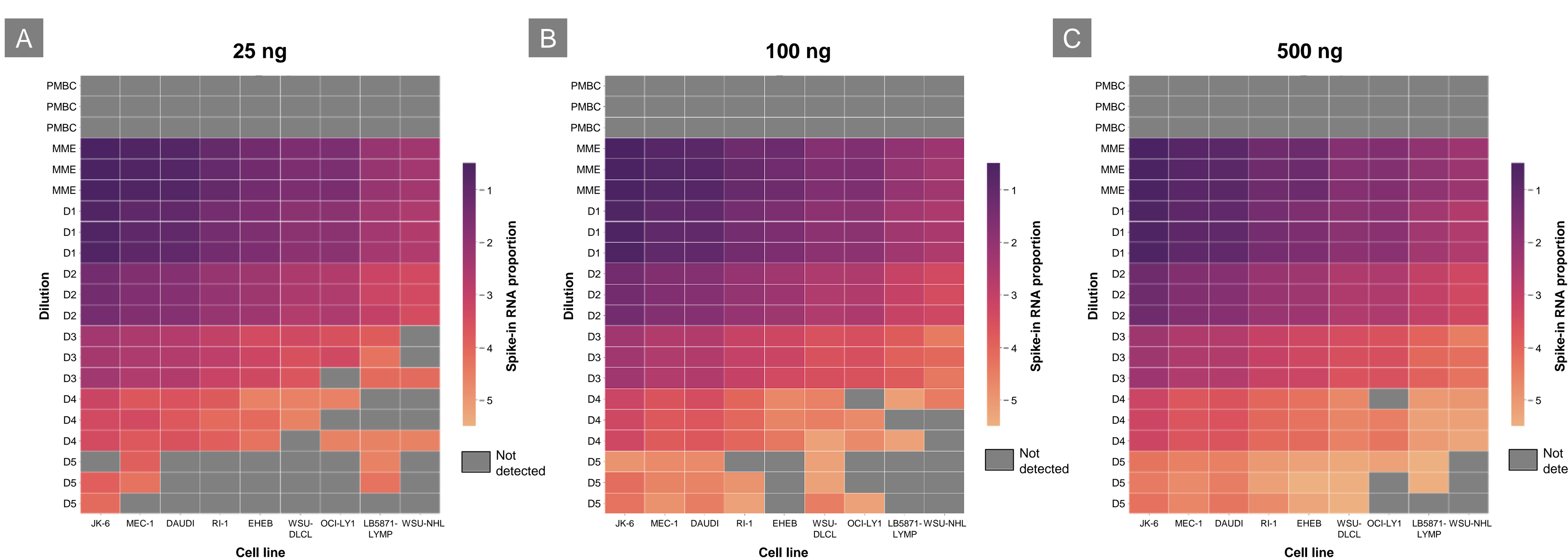



**Figure 8. Sensitivity and accuracy of SMART-Seq Human BCR (with UMIs).** The limits of detection for specific heavy chain clonotypes corresponding to cells lines in the MME spike-in cell line mix. Samples correspond to the serial dilutions outlined in Figure 1, Panel B. **Panel A.** 25 ng input libraries. **Panel B.** 100 ng input libraries. **Panel C.** 500 ng input libraries. Clonotypes corresponding to the spike-in cell lines were not detected in PBMC RNA samples at all three input concentrations, demonstrating accuracy.

## 9 Performance across Illumina instruments

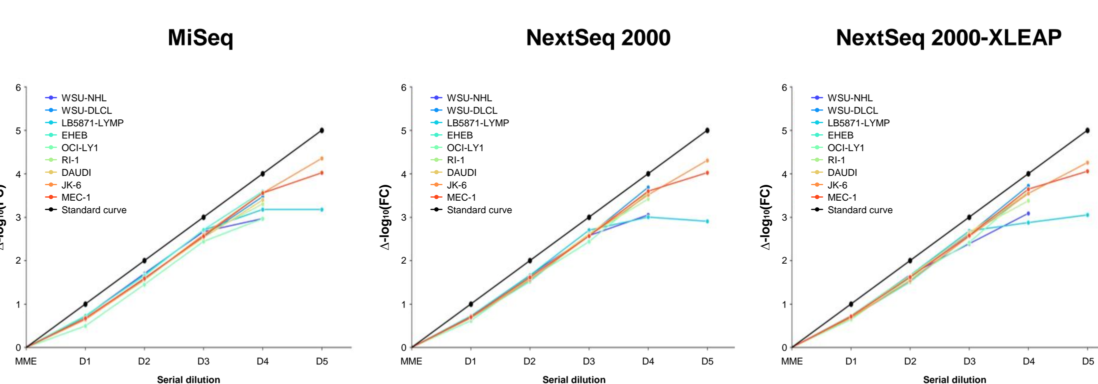



**Figure 9. Alignment statistics for 25 ng input libraries across different sequencing instruments.** Impact of sequencing platforms on detection of clonotypes corresponding to spike-in cell lines in MME cell mix. The serial dilutions on the x-axis corresponds serial dilutions of the MME cell mix as outlined in Figure 1, Panel B.

## Conclusions

- Libraries produced with SMART-Seq Human BCR (with UMIs) show an exponential increase in isotype count corresponding to an increase in RNA input, with high reproducibility between technical replicates.
- Correlation between the ddPCR assay data and NGS data shows that SMART-Seq Human BCR (with UMIs) is free from potential biases introduced by PCR, primers, or sequencing instruments.
- The linearity of detection of spike-in clonotypes at various dilutions demonstrates the accuracy of SMART-Seq Human BCR (with UMIs).
- SMART-Seq Human BCR (with UMIs) kit has outstanding sensitivity, as the limit of detection for tested spike-in concentrations is $\leq 10^{-6}$.
- SMART-Seq Human BCR (with UMIs) displays excellent sensitivity and linearity regardless of which sequencing platform is used.

**Takara Bio USA, Inc.**
United States/Canada: +1.800.662.2566 • Asia Pacific: +1.650.919.7300 • Europe: +33.(0)1.3904.6880 • Japan: +81.(0)77.565.6999
FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES. © 2024 Takara Bio Inc. All Rights Reserved. All trademarks are the property of Takara Bio Inc. or its affiliate(s) in the U.S. and/or other countries or their respective owners. Certain trademarks may not be registered in all jurisdictions. Additional product, intellectual property, and restricted use information is available at takarabio.com

Meeting details and poster download:
takarabio.com/AIRR

800.662.2566
takarabio.com